

Amendment of the Claims

1. (Previously Presented) A method of generating an index for a sequence that supports a non-contiguous subsequence match, comprising:

receiving a sequence;

receiving a window size;

encoding the sequence into a weighted-sequence;

encoding the weighted sequence into one or more one-dimensional sequences, wherein the length of each of the one or more one-dimensional sequences is less than the window size;

inserting each of the one or more one-dimensional sequences as one or more trie nodes into a trie structure; and

generating an index, wherein generating the index comprises:

generating a current sequential ID and a maximum sequential ID pair

for each of the one or more trie nodes, wherein the current sequential ID of any descendant of a given trie node of the one or more trie nodes is between the current sequential ID of the given trie node and the maximum sequential ID;

generating an iso-depth link for each unique symbol in each of the one or more one-dimensional sequences, wherein the iso-depth link comprises trie nodes under the symbol; and

generating an offset list comprising an original position for each of one or more subsequences in the weighted-sequence.

2. (Previously Presented) The method of claim 1, wherein encoding the

sequence into the weighted-sequence comprises encoding the sequence with weights represented by real numbers.

3. (Previously Presented) The method of claim 2, wherein encoding the sequence with weights represented by real numbers comprises discretizing the sequence into a number of equi-width units.

4. (Previously Presented) The method of claim 1, wherein inserting each of the one or more one-dimensional sequences into the trie structure is performed using a depth-first traversal.

5. (Previously Presented) The method of claim 1, further comprising creating a weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a one-dimensional buffer.

6. (Previously Presented) The method of claim 1, further comprising creating a weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a B⁺ tree.

7. (Previously Presented) The method of claim 1, further comprising creating a weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a linked list.

8. (Previously Presented) The method of claim 1, wherein receiving the sequence comprises receiving one or more elements in the sequence, wherein each of the one or more elements are represented by one or more pairs of symbol and weight elements.

9. (Previously Presented) The method of claim 8, the symbol elements correspond to a non-uniform frequency distribution.

10. (Previously Presented) The method of claim 9, further comprising reordering the one or more one-dimensional sequences using the non-uniform frequency distribution to generate a new sequence prior to inserting each of the one or more one-dimensional sequences into the trie structure.

11. (Currently Amended) The method of claim 10, wherein reordering the one or more one-dimensional sequences using the non-uniform frequency distribution to generate a new sequence prior to inserting each of the one or more one-dimensional sequences into the trie structure, comprises:

(a) adding an offset $2*w*r$ to each of the weight elements to generate a new weight, wherein w is a window size, and r is a frequency rank for a symbol of each of the symbol elements;

(b) sorting the pairs of symbol and weight elements by the new weight;

(c) placing a moving window of size $2*w*A$ on the new sequence, wherein A is a ~~frequency~~total number of the symbols; and

(d) indexing the new sequence in a new window.

12. (Previously Presented) The method of claim 1, wherein receiving the sequence comprises receiving one or more scientific datasets, transforming each of the one or more scientific datasets into one or more sequences, and concatenating the one or more sequences to form a long sequence.

13. (Cancelled)

14. (Cancelled)

15. (Currently Amended) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform ~~instructions executable by the machine to perform a~~ method steps of generating an index for a sequence that supports a non-contiguous subsequence match, the method steps comprising:

receiving a sequence;

receiving a window size;

encoding the sequence into a weighted-sequence;

encoding the weighted sequence into one or more one-dimensional sequences,

wherein the length of each of the one or more one-dimensional sequences is less than the window size;

inserting each of the one or more one-dimensional sequences as one or more trie nodes into a trie structure; and

generating an index, wherein generating the index comprises:

generating a current sequential ID and a maximum sequential ID pair for each of the one or more trie nodes, wherein the current sequential ID of any descendant of a given trie node of the one or more trie nodes is between the current sequential ID of the given trie node and the maximum sequential ID;

generating an iso-depth link for each unique symbol in each of the one or more one-dimensional sequences, wherein the iso-depth link comprises trie nodes under the symbol; and

generating an offset list comprising an original position ~~for~~^{of} each of the one or more subsequences in the weighted-sequence.

16. (Cancelled)